Quaderni di Statistica e Matematica
applicata alle Scienze Economico-Sociali
VOLUME IV DICEMBRE 1981

SUI MAGGIORANTI DELLA DISTRIBUZIONE ASINTOTICA DI UN INDICE DI CONCENTRAZIONE FRA LE CLASSI DI GINI*

Michele ZENGA - Antonio BRUNAZZO**

0. Sommario

Nelle indagini socio-economiche, spesso si hanno problemi che riguardano la difforme distribuzione di una quantità trasferibile fra le k classi in cui può essere divisa una popolazione. Una misura della concentrazione fra le classi è data dal noto indice di concentrazione di Gini:

$$\Delta = \frac{D_r}{2\bar{X}}$$

in cui D_r indica la differenza media assoluta con ripetizione fra le medie delle classi e $\bar{X}$ indica, come al solito, la media artimetica generale della quantità trasferibile. Con semplici passaggi è facile esprimere l'indice di concentrazione fra le classi anche nel seguente modo:

$$\Delta = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} | p_i q_j - p_j q_i|$$

in cui:

i) p_i indica la frequenza relativa delle unità appartenenti alla classe $i^{\underline{ma}}$;

ii) q_i indica la quota del carattere trasferibile spettante alla classe $i^{\underline{ma}}$.

Nell'ipotesi in cui sono noti i valori di q_i della popolazione e si disponga di un campione bernoulliano che permetta di stimare le frequenze relative p_i

---

* Lavoro presentato, in forma ristretta, alla XXXI riunione scientifica della Società Italiana di Statistica, Torino 5-7 aprile 1982. I capitoli 0,1,2,3 sono dovuti a M. Zenga mentre i capitoli 4,5,6 e 7 sono dovuti a A. Brunazzo.

** Facoltà di Economia e Commercio, Libera Università degli Studi di Trento

della popolazione attraverso le frequenze $\hat{p}_i$ del campione, è possibile misurare la concentrazione campionaria fra le classi con la statistica:

$$G = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} | \hat{p}_i q_j - \hat{p}_j q_i |$$

La distribuzione asintotica di G è stata ricavata da Colombi [1] nel caso $\Delta = 0$ e $k = 3$ e da Brunazzo [2] nel caso $\Delta = 0$, $k = 4$ e $q_1 = q_2 = q_3 = q_4 = \frac{1}{4}$.

In questo lavoro, sempre per l'ipotesi $\Delta = 0$, si ricavano con due diverse procedure, valori maggioranti dei quantili di G. La prima procedura utilizza la classica disuguaglianza di Bonferroni, l'altra utilizza la distribuzione della v.c. Chi-quadrato centrale.

Le due procedure, in qualche modo, si compensano nel senso che, per valori piccoli di k, i maggioranti che si ricavano con la disuguaglianza di Bonferroni sono minori degli altri ed all'aumentare di k le posizioni si invertono.

Il grado di approssimazione fra i maggioranti qui proposti ed i valori esatti dei quantili trovati da Colombi e da Brunazzo per il caso $k = 3$ e $k = 4$ è del tutto soddisfacente e giustifica l'impiego dei maggioranti in problemi di inferenza statistica.

1. Introduzione e simbologia

Si indichi con X l'intensità di un carattere trasferibile di cui si vuole misurare la concentrazione fra le k classi in cui è divisa la popolazione. Sia N_i il numero delle unità appartenenti alla classe $i^{\underline{ma}}$ $(i = 1,2,\ldots, k)$ e sia X_{ij} l'intensità della generica $j^{\underline{ma}}$ unità della $i^{\underline{ma}}$ classe $(j = 1,2,\ldots, N_i)$.

E' possibile, così calcolare le seguenti grandezze statistiche:

i) $$p_i = \frac{N_i}{\sum_{i=1}^{k} N_i} = \frac{N_i}{N}\,,$$ frequenza relativa delle unità appartenenti alla classe $i^{\underline{ma}}$.

ii) $\bar{X}_i = \dfrac{\sum_{j=1}^{N_i} X_{ij}}{N_i}$, media aritmetica del carattere nella classe $i^{\underline{ma}}$;

iii) $\bar{X} = \dfrac{\sum_{i=1}^{k} N_i \bar{X}_i}{\sum_{i=1}^{k} N_i}$, media artimetica generale;

iV) $Q_i = N_i \bar{X}_i$, quantità del carattere appartenente alla classe $i^{\underline{ma}}$;

V) $q_i = \dfrac{Q_i}{\sum_{i=1}^{k} Q_i}$

$= \dfrac{N_i \cdot \bar{X}_i}{N\,\bar{X}}$ quota del carattere appartenente alla classe $i^{\underline{ma}}$.

Per misurare la concentrazione fra le classi si può usare l'indice di concentrazione di Gini:

$$\Delta = \frac{D_r}{2\bar{X}}$$

in cui D_r indica la differenza media assoluta con ripetizione fra le medie $\bar{X}_i$ delle classi. La concentrazione fra le classi, si può, esplicitando l'espressione di D_r, riscrivere come segue:

$$\Delta = \frac{1}{2\bar{X}} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| \bar{X}_j - \bar{X}_i \right| = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| \frac{\bar{X}_j}{\bar{X}} - \frac{\bar{X}_i}{\bar{X}} \right| p_i p_j \;.$$

Tenuto presente che $\dfrac{\bar{X}_i}{\bar{X}} = \dfrac{q_i}{p_i}$, l'espressione precedente diventa:

$$\Delta = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| p_i q_j - p_j q_i \right| \;. \tag{1.1}$$

Nell'ipotesi di equiripartizione si ha:

$$\Delta = 0 \leftrightarrow p_i = q_i \quad (i = 1,2,\ldots, k).$$

Si estragga, con riposizione, dalla popolazione un campione casuale di ampiezza n e si indichi n_i il numero delle unità estratte appartenenti alla classe $i^{\underline{ma}}$ (ovviamente: $n_i > 0; i = 1,2,\ldots, k; \sum n_i = n$). Le frequenze $(n_1, n_2, \ldots n_k)$ costituiscono la determinazione campionaria di una distribuzione multinomiale ' $(n; p_1, p_2, \ldots, p_k)$. Nell'ipotesi in cui siano noti i valori q_i della popolazione, dai dati campionari è possibile calcolare l'indice di Gini che assumerà la forma:

$$G = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| \frac{n_i}{n} q_j - \frac{n_j}{n} q_i \right| \qquad (1.2)$$

Nei paragrafi che seguono si indagherà sulla distribuzione campionaria della (1.2) nell'ipotesi che $\Delta = 0$.

2. Impiego della disuguaglianza di Bonferroni, caso k = 3

In questo paragrafo verrá esaminato, in dettaglio, il procedimento che ha permesso di ricavare i maggioranti con l'uso della disuguaglianza di Bonferroni. Si è preferito esaminare a parte il caso k = 3 in quanto la trattazione per un k generico risulta abbastanza complessa, come si noterà nel paragrafo 4 e seguenti. Con semplici passaggi è possibile descrivere la (1.2) come segue:

$$G = \frac{1}{n} \left\{ q_1 q_2 \left| \frac{n_1}{q_1} - \frac{n_2}{q_2} \right| + q_1 q_3 \left| \frac{n_1}{q_1} - \frac{n_3}{q_3} \right| + q_2 q_3 \left| \frac{n_2}{q_2} - \frac{n_2}{q_3} \right| \right\}. \qquad (2.1)$$

Si considerino ora le seguenti sei regioni R_i, $(i = 1,2,\ldots, 3!)$ dello spazio R^2:

$$R_1 = \left\{ (n_1, n_2, n_3) : \frac{n_3}{q_3} \leqslant \frac{n_2}{q_2} \leqslant \frac{n_1}{q_1} \right\}$$

$$R_2 = \left\{ (n_1, n_2, n_3) : \frac{n_1}{q_1} \leqslant \frac{n_2}{q_2} \leqslant \frac{n_3}{q_3} \right\}$$

$$R_3 = \left\{ (n_1, n_2, n_3) : \frac{n_3}{q_3} \leqslant \frac{n_1}{q_1} \leqslant \frac{n_2}{q_2} \right\}$$

$$R_4 = \left\{ (n_1, n_2, n_3) : \frac{n_2}{q_2} \leqslant \frac{n_1}{q_1} \leqslant \frac{n_3}{q_3} \right\}$$

$$R_5 = \left\{(n_1, n_2, n_3) : \frac{n_2}{q_2} \leqslant \frac{n_3}{q_3} \leqslant \frac{n_1}{q_1}\right\}$$

$$R_6 = \left\{(n_1, n_2, n_3) : \frac{n_1}{q_1} \leqslant \frac{n_3}{q_3} \leqslant \frac{n_2}{q_2}\right\}$$

Si indichi con G_i l'espressione con cui si può scrivere la (2.1), senza ricorrere ai valori assoluti, nell'ipotesi $(n_1, n_2, n_3) \in R_i$.

In particolare si ha:

$$G_1 = \frac{1}{n}\left\{q_1 q_2 \left(\frac{n_1}{q_1} - \frac{n_2}{q_2}\right) + q_i q_3 \left(\frac{n_1}{q_1} - \frac{n_3}{q_3}\right) + q_2 q_3 \left(\frac{n_2}{q_2} - \frac{n_3}{q_3}\right)\right\};$$

$$G_2 = -G_1;$$

$$G_3 = \frac{1}{n}\left\{-q_1 q_2 \left(\frac{n_1}{q_2} - \frac{n_2}{q_2}\right) + q_1 q_3 \left(\frac{n_1}{q_1} - \frac{n_3}{q_3}\right) + q_2 q_3 \left(\frac{n_2}{q_2} - \frac{n_3}{q_3}\right)\right\};$$

$$G_4 = -G_3;$$

$$G_5 = \frac{1}{n}\left\{q_1 q_2 \left(\frac{n_1}{q_1} - \frac{n_2}{q_2}\right) + q_1 q_3 \left(\frac{n_1}{q_1} - \frac{n_3}{q_3}\right) - q_2 q_3 \left(\frac{n_2}{q_2} - \frac{n_3}{q_3}\right)\right\};$$

$$G_6 = -G_5.$$

Riordinando i termini delle espressioni precedenti e ricordando che $n_3 = n - n_1 - n_2$, si ha ancora:

$$G_1 = \frac{1}{n}\left\{n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3)\right\}; \qquad (2.2)$$

$$G_3 = \frac{1}{n}\left\{n_1 (1 - q_2) + n_2 (1 + q_1) - n (1 - q_3)\right\}; \qquad (2.3)$$

$$G_5 = \frac{1}{n}\left\{n_1 (1 - q_2) - n_2 (1 - q_1) - n (q_1 - q_2)\right\}. \qquad (2.4)$$

Ovviamente è possibile considerare la G_i anche fuori dalla regione R_i, però il suo valore, in questo caso, non coincide con quello di G.
Si consideri ora il luogo dei punti dello spazio campionario R^2 in cui $G_1 = g$, $(g > 0)$: trattasi dei punti caratterizzati dalla retta di equazione:

$$n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3) = ng.$$